

# PROPIEDADES PSICOMÉTRICAS DEL CUESTIONARIO SAHLSA-50, PRUEBA ESTANDARIZADA PARA MEDIR EL CONOCIMIENTO SOBRE LA SALUD

## Psychometric properties of the SAHLSA-50, a standardized test to evaluate the health literacy

Paola Padilla-Santoyo<sup>1</sup>, Carlos Vilchez-Román<sup>2</sup>

### RESUMEN

**Objetivo:** analizar las propiedades psicométricas de una prueba que mide el conocimiento sobre la salud, la cual fue diseñada originalmente para pacientes que hablan español.

**Material y método:** prueba SAHLSA-50 aplicada en una muestra de 49 mujeres en edad fértil, usuarias del Servicio Ginecológico del Hospital Dos de Mayo. Para el análisis de datos empleamos el coeficiente alfa de Cronbach y alfa si el ítem es borrado.

**Resultados:** el puntaje SAHLSA-50 logró una distribución casi normal y un nivel moderadamente alto de confiabilidad ( $\alpha=0,8773$ ); sin embargo, identificamos un grupo de ítems que deben ser revisados debido a su efecto sobre la confiabilidad de la escala.

**Conclusiones:** para saber si los pacientes de los servicios de salud realmente entienden el material informativo y promocional que reciben, es necesario realizar estudios de este tipo, pero siempre verificando si el instrumento de recolección de datos tiene las propiedades psicométricas adecuadas.

**Palabras clave:** reproducibilidad de resultados, validez de las pruebas.

### ABSTRACT

**Objective:** To analyze the psychometric properties of a test for measure the health literacy, which originally was designed for spanish speaking patients.

**Material and method:** SAHLSA-50 test with a sample of 49 women in reproductive age, users of the gynaecological service at the Dos de Mayo hospital, in Lima (Perú). For the data analysis we used the the Cronbach's alpha coefficient and the value of alpha if item is deleted.

**Results:** The SAHLSA-50 score had an almost normal distribution and a moderate high level of reliability ( $\alpha=0.8773$ ), but we identified a group of items that should be reviewed because its effect on the reliability of the scale.

**Conclusions:** To find out if patients of health services actually understand the promotional and informative material they receive it is necessary to carry out this kind of studies, but always analyzing if the data collection tool has adequate psychometric properties.

**Keywords:** Reproducibility of Results, Validity of Tests

### INTRODUCCIÓN

Los profesionales de la salud que usan diferentes instrumentos de medición no sólo desean saber si estos están disponibles para ser usados, es decir operativos y en buenas condiciones; también les preocupa conocer hasta qué punto obtienen buenas mediciones y si de verdad miden lo que deberían medir. En síntesis, les interesa conocer qué tan válidos y confiables son sus instrumentos de medición, no sólo termómetros o

electroencefalogramas, sino también pruebas de ansiedad, de personalidad, así como aquellas que miden el conocimiento sobre la salud.

El conocimiento sobre la salud es un concepto que puede ser abordado desde tres frentes: el cuidado de la salud, la educación y los medios informativos <sup>1, 2</sup>. Los investigadores lo definen como la capacidad para obtener, procesar y comprender la información, y los servicios necesarios para tomar adecuadas decisiones sobre el cuidado de la salud <sup>1,2</sup>; es decir, no se limita a la comprensión de folletos, afiches y documentos similares, sino que tiene que ver con la comprensión del alcance de los servicios de salud <sup>3,4</sup>. La manera más práctica de determinar el conocimiento sobre la salud en la población es emplear pruebas o cuestionarios estandarizados.

<sup>1</sup> Bibliotecóloga del Colegio Peruano Norteamericano Abraham Lincoln.

<sup>2</sup> Licenciado en Bibliotecología y Ciencias de la Información. Docente en la Maestría de Bibliotecología (Universidad Nacional Mayor de San Marcos).

### 1.1. Pruebas estandarizadas

Entre las pruebas que miden el conocimiento sobre la salud, las más difundidas han sido dos herramientas desarrolladas inicialmente para pacientes de habla inglesa<sup>5, 6, 3, 1, 7, 8, 9</sup>: estimado rápido de idioma adulto en medicina (REALM, por sus siglas en inglés), y prueba de conocimiento funcional sobre salud en adultos (TOFHLA, por sus siglas en inglés).

Aun cuando suelen ser consideradas dentro del mismo grupo, en realidad miden habilidades distintas. La primera está enfocada en la capacidad para pronunciar correctamente un conjunto de términos médicos; en ese sentido, es más una prueba de reconocimiento lingüístico<sup>8</sup>. Por su parte, la prueba TOFHLA mide la habilidad para asociar términos referidos al cuidado de la salud con otros similares; es decir, hasta qué punto los entrevistados son capaces de identificar el término correcto dentro de un grupo de términos relacionados<sup>5</sup>.

De las dos pruebas antes mencionadas, el cuestionario TOFHLA es el que mejor encaja en el concepto de conocimiento sobre la salud.

### 1.2. Limitaciones de las pruebas estandarizadas

Si bien es cierto, ambos cuestionarios cuentan con versión en español, cuando pacientes que hablan inglés y español los responden, suelen obtener mejores puntajes en la versión en inglés<sup>5,8</sup>. Ello revela que al menos con estas pruebas, los investigadores obtienen mayor utilidad cuando las utilizan en su idioma original.

De otro lado, tenemos el problema de la duración, ya que para su aplicación TOFHLA toma no menos de 12 minutos<sup>5</sup>; por esa razón, los creadores del TOFHLA desarrollaron una versión breve que toma entre 6 y 10 minutos en ser contestada<sup>5,8</sup>.

### 1.3. Propiedades psicométricas de las pruebas estandarizadas

Existen dos características que deben tener los instrumentos de recolección de datos: confiabilidad y validez. La primera alude a la estabilidad en la medición de los atributos, mientras que la validez permite determinar si los resultados obtenidos se adecuan a los supuestos teóricos en los que se basa el instrumento de recolección. En realidad, existen distintos tipos de validez, aunque los más empleados son la validez predictiva, la validez concurrente y la validez de constructo.

En general, por el tipo de técnicas de análisis empleadas, el estudio de las propiedades psicométricas requiere trabajar con muestras grandes ( $n > 300$ ). Por ejemplo, con relación a la validez de constructo, la técnica multivariada empleada suele ser el análisis de componentes principales.

El análisis de la confiabilidad puede abordarse a través de dos enfoques: la teoría clásica de los test (TCT) o la teoría de respuesta al ítem (TRI). La primera es usada en la mayoría de análisis psicométricos; sin embargo, presenta algunas limitaciones: las características de la prueba y las puntuaciones individuales no pueden ser separadas; el puntaje obtenido asume un nivel de complejidad similar para todos los participantes, entre otras. Además, hoy se sabe que la probabilidad de responder con éxito las preguntas de una prueba depende de atributos personales tales como la pericia y los conocimientos previos<sup>10,11</sup>; por ello ha sido necesario desarrollar enfoques alternativos.

Precisamente, la teoría de respuesta al ítem (TRI) surgió como una alternativa para solucionar las críticas a la teoría clásica de los tests (TCT); en ese sentido la TRI no solo considera el nivel de complejidad de cada ítem y la capacidad para diferenciar entre puntajes altos y bajos, sino que también toma en cuenta la probabilidad que el participante haya adivinado la respuesta correcta<sup>10,12</sup>. Ciertamente, es un modo de análisis más completo y detallado, pero por la misma razón requiere muestras grandes ( $n > 300$ ) y modelos estadísticos más avanzados, que incluyen el uso de ecuaciones de regresión logística con dos y tres parámetros.

El propósito de esta investigación es analizar las propiedades psicométricas de una prueba que mide el conocimiento sobre la salud, la cual fue diseñada originalmente para pacientes que hablan español.

## MATERIAL Y MÉTODO

### Población

Participaron 49 mujeres en edad reproductiva, quienes son usuarias del Servicio Ginecológico del Hospital Dos de Mayo (Lima, Perú). Al momento del estudio, las encuestadas tenían entre 22 y 50 años. Antes de responder el cuestionario, cada participante recibió una breve explicación sobre la naturaleza del estudio y la confidencialidad con que se manejarían los datos.

## Instrumento de recolección de datos

El cuestionario SAHLSA-50 (*Short Assessment of Health Literacy in Spanish-speaking adults*) permite identificar el nivel de conocimiento sobre la salud. La prueba tiene 50 ítems agrupados en dos secciones: una para evaluar la correcta pronunciación del término y la otra para comprobar la capacidad de asociar los ítem propuestos; para ello los participantes deben elegir una de las tres opciones: el término correcto, un distractor y la opción "No sé". El formato ha sido diseñado de tal manera que las personas encuestadas puedan responder las preguntas directamente, tomando entre 6 y 10 minutos para completarlo. Para obtener el puntaje SAHLSA-50 de cada participante sumamos sus aciertos <sup>8</sup>.

## Técnicas de análisis de datos

Debido al tamaño de la muestra ( $n=49$ ), el análisis de las propiedades psicométricas estuvo centrado en el análisis de confiabilidad, para lo cual empleamos la TCT; sin embargo, siempre que el tamaño de la muestra y los conocimientos de los investigadores lo permitan, el análisis de confiabilidad debe basarse en la TRI, debido a las limitaciones ya señaladas de la TCT. Los

investigadores han planificado una segunda aplicación de la prueba SAHLSA-50 con una muestra mediana ( $n>200$ ), la cual permitirá emplear los postulados de la TRI.

Identificamos el coeficiente alfa de Cronbach de toda la prueba y luego lo comparamos con el valor del alfa si el ítem es borrado. Posteriormente, determinamos los ítem que deben ser reformulados o eliminados en la próxima aplicación del cuestionario SAHLSA-50.

**Hipótesis:** existe un grupo de ítems que maximiza la confiabilidad de los puntajes del cuestionario SAHLSA-50.

## RESULTADOS

### Datos demográficos

Las 49 mujeres que participaron en el estudio obtuvieron un puntaje promedio de  $37 \pm 7,15$  puntos (ver figura 1). Los investigadores han encontrado que este puntaje depende del total de años estudiados y del estado civil de las encuestadas <sup>11</sup>.

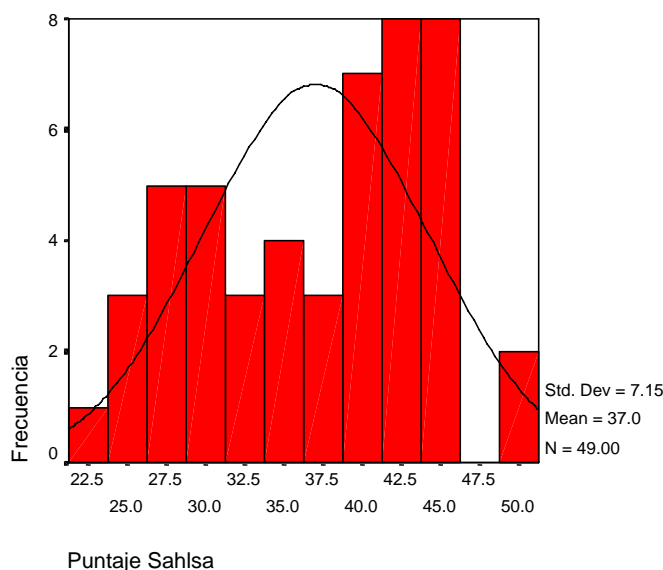


Figura1: distribución de frecuencia del puntaje SAHLSA-50

### Análisis de confiabilidad

El valor del alfa de Cronbach ( $\alpha=0,8773$ ) reveló un nivel de confiabilidad moderadamente alto en esta aplicación de la prueba SAHLSA-50; sin embargo detectamos

algunos ítems que afectaron la confiabilidad del cuestionario a nivel de cada ítem (ver tabla 1).

La revisión detallada de aquellos ítems que al ser eliminados de la escala produjeron una mejora en la confiabilidad de la prueba (los ítems *empleo, comidas, pastilla, aborto, ataque, nutrición, alérgico, calorías*), llevó

a los investigadores a concluir que se trata de ítems bastante sencillos, en los que la probabilidad de acierto está casi garantizada por lo que no son muy efectivos para discriminar y determinar que tan alto o bajo conocimiento sobre la salud tienen las personas que participaron del estudio.

En ese sentido, encontramos que efectivamente existe un grupo de ítem que maximiza la confiabilidad del

puntaje de la prueba SAHLSA-50. Si bien es cierto los diseñadores de la prueba SAHLSA-50 emplearon la TRI para seleccionar los ítem que permitieron mejorar las propiedades psicométricas del cuestionario, en esta versión de SAHLSA-50 no fue posible aplicarla debido a las restricciones por el tamaño de la muestra.

**Tabla 1: confiabilidad de los ítems del cuestionario SAHLSA-50**

	Promedio de la escala si el ítem es borrado	Varianza de la escala si el ítem es borrado	Correlación corregida ítem-total	Alfa si el ítem es borrado
ÍTEM01	36,6122	48,3257	,3021	,8759
<b>ÍTEM02</b>	<b>36,2245</b>	<b>50,6361</b>	<b>,0000</b>	<b>,8776</b>
ÍTEM03	36,2857	49,7500	,2421	,8763
ÍTEM04	36,3061	50,0085	,1408	,8774
ÍTEM05	36,3061	49,1335	,3677	,8748
<b>ÍTEM06</b>	<b>36,2449</b>	<b>50,8971</b>	<b>-,1381</b>	<b>,8788</b>
ÍTEM07	36,3061	49,0085	,4004	,8745
ÍTEM08	36,2857	49,3333	,3656	,8751
ÍTEM09	36,2653	49,6990	,3183	,8757
ÍTEM10	36,5714	48,7917	,2401	,8770
ÍTEM11	36,4490	47,5026	,5086	,8720
ÍTEM12	36,4694	47,9626	,4129	,8737
ÍTEM13	36,4490	47,4192	,5234	,8717
ÍTEM14	36,2449	50,0221	,2937	,8763
ÍTEM15	36,7755	48,6777	,2433	,8772
<b>ÍTEM16</b>	<b>36,2653</b>	<b>50,3240</b>	<b>,0959</b>	<b>,8775</b>
ÍTEM17	36,6735	44,8078	,8287	,8647
ÍTEM18	36,5714	47,1250	,4967	,8720
ÍTEM19	36,3061	48,7168	,4772	,8736
ÍTEM20	36,5510	47,8359	,3930	,8740
ÍTEM21	36,4082	49,2466	,2252	,8768
ÍTEM22	36,2653	50,1990	,1402	,8772
<b>ÍTEM23</b>	<b>36,2245</b>	<b>50,6361</b>	<b>,0000</b>	<b>,8776</b>
ÍTEM24	36,9184	48,4932	,2970	,8759
ÍTEM25	36,3878	48,4923	,3854	,8742
ÍTEM26	37,0612	49,3503	,2185	,8768
ÍTEM27	36,6122	48,3257	,3021	,8759
<b>ÍTEM28</b>	<b>36,3469</b>	<b>50,2313</b>	<b>,0629</b>	<b>,787</b>
ÍTEM29	36,3061	49,2168	,3459	,751
ÍTEM30	36,6531	46,1063	,6303	,691
ÍTEM31	36,3878	48,6173	,3609	,746
<b>ÍTEM32</b>	<b>36,2653</b>	<b>50,6156</b>	<b>-,0069</b>	<b>,784</b>

ÍTEM33	36,3061	48,9668	,4114	,743
ÍTEM34	37,0204	47,8954	,4568	,730
ÍTEM35	36,3469	48,6063	,4158	,739
ÍTEM36	36,5306	47,3793	,4742	,724
ÍTEM37	36,7755	48,5944	,2554	,769
<b>ÍTEM38</b>	<b>36,4898</b>	<b>49,2968</b>	<b>,1820</b>	<b>,779</b>
ÍTEM39	36,2857	49,4167	,408	,753
<b>ÍTEM40</b>	<b>36,8776</b>	<b>50,9847</b>	<b>-,844</b>	<b>,832</b>
ÍTEM41	36,3061	49,2585	,350	,752
ÍTEM42	36,2653	50,0740	,845	,768
ÍTEM43	36,2857	49,7083	,544	,762
ÍTEM44	36,8571	48,4583	,862	,762
ÍTEM45	36,8571	46,3750	,066	,697
ÍTEM46	36,4082	48,2883	,036	,739
ÍTEM47	36,6531	46,2313	,111	,695
ÍTEM48	36,7143	46,2500	,013	,697
ÍTEM49	36,4898	48,4634	,178	,754
ÍTEM50	36,5306	47,8793	,941	,740

## DISCUSIÓN

Para analizar las propiedades psicométricas de todo instrumento de medición es importante trabajar con muestras grandes que permitan emplear técnicas como el análisis de componentes principales o enfoques teóricos como la teoría de respuesta al ítem (TRI). Sin embargo, debido a que esta investigación no contó con financiamiento (más allá del generoso aporte de las personas que aplicaron la encuesta y los limitados recursos invertidos por los investigadores para reproducir el formato de la prueba), tuvimos limitaciones al momento de analizar los datos. Por ello, no pudimos determinar la confiabilidad de los ítems empleando la teoría de respuesta al ítem (TRI), tal como lo hicieron los diseñadores de la prueba SAHLA-50<sup>8</sup>. Las restricciones ya señaladas únicamente nos permitieron trabajar con la teoría clásica de los test (TCT). En ese sentido, el criterio para seleccionar los ítems que maximizan la confiabilidad de la escala se basó en el grado de variación de los puntajes y no en la probabilidad de ocurrencia ni en la complejidad de los ítems<sup>8</sup>. Con relación a los resultados del análisis de confiabilidad, constatamos que a pesar del cuidado que los creadores de instrumentos de recolección de datos tienen al momento de crearlos, siempre debemos verificar si la herramienta alcanzó los niveles de validez y confiabilidad esperados. En el caso de la prueba SAHLA-50, comprobamos que -a pesar de haber sido diseñada para pacientes que hablan español<sup>8</sup> - su aplicación en un

hospital de la ciudad de Lima revela que es necesario revisar ocho de los 50 ítems de la prueba antes de su aplicación a mayor escala. Como hemos señalado en otro estudio<sup>11</sup>, para que las campañas de salud pública -emprendidas por el Estado peruano a través del Ministerio de Salud y las diferentes direcciones de salud de los gobiernos regionales- sean efectivas, es necesario determinar hasta qué punto el material informativo y promocional efectivamente es comprendido por los supuestos beneficiarios. En el cuestionario SAHLA-50 encontramos ocho ítems (empleo, comidas, pastilla, aborto, ataque, nutrición, alérgico y calorías) de uso común en nuestro medio; si revisamos la definición de algunos términos como comida y empleo en el Diccionario de la Real Lengua Española, encontraremos que la palabra *comida* es definida como "alimento que se toma al mediodía o primeras horas de la tarde" y el término *empleo* como "ocupación u oficio". Es decir, se trata de términos que no necesariamente reflejan el conocimiento de los pacientes en temas de salud, por lo que recomendamos eliminar ambos ítems del cuestionario o caso contrario reemplazarlos por términos más pertinentes. De allí la importancia de realizar un estudio previo de los ítems a utilizar, considerando principalmente el contexto local donde se llevará a cabo la investigación. Sugerimos realizar estudios de este tipo utilizando la herramienta SAHLA-50 en las diferentes postas médicas, hospitales, clínicas u otros centros de atención, con la finalidad de detectar si

existe analfabetismo funcional en los pacientes a quienes se les brindan servicios de salud. En caso los estudios evidencien bajo nivel de alfabetismo en salud, podremos determinar que existe una falta de comprensión del material que se proporciona en estas áreas (ya sean recetas médicas, folletos, afiches, materiales preventivos, etc.), por lo que se deberá tomar acciones tales como realizar charlas o difundir material informativo con un lenguaje más fácil de comprender por los pacientes. Por ser la primera investigación sobre alfabetismo en salud empleando instrumentos estandarizados, no hay muchas lecciones aprendidas que compartir, pero estamos seguros que en el futuro esta situación cambiará ya que contaremos con las experiencias de otros investigadores. Asimismo, es importante señalar que la presencia de diversos profesionales como bibliotecólogos, comunicadores, educadores y especialistas en ciencias de la salud cumple un papel fundamental en el desarrollo de este tipo de estudios: los bibliotecólogos llevando a cabo las mediciones correspondientes que permitan determinar el nivel de alfabetismo en salud, los comunicadores dando una estructura adecuada a los contenidos del material informativo, los educadores aplicando programas de alfabetización y finalmente los especialistas en la ciencia de la salud aportando con el conocimiento médico. Para una mejor aplicación de esta herramienta recomendamos trabajar con muestras más grandes, porque de esta forma podremos analizar con mayor profundidad la validez y confiabilidad del instrumento; por ejemplo, verificando si la probabilidad de responder con éxito a las preguntas de la prueba depende de la experiencia personal y los conocimientos previos, tal como lo han revelado otros estudios <sup>10, 11</sup> o si depende de otros factores. Asimismo, es de suma importancia identificar aquellos ítems que evidencian, según el contexto en el cual nos encontremos, una elemental asociación entre términos referidos al cuidado de la salud, ya que esto conllevará a que la confiabilidad de la prueba no se vea afectada y logremos una mejor medición del alfabetismo en salud de los participantes.

## 5. Agradecimientos

Agradecemos la colaboración de los doctores Jazmín Reyes y Pedro Bustíos, quienes brindaron las facilidades del caso para llevar adelante el presente estudio; y de la

señorita Rocío Quiliano Terreros por las coordinaciones realizadas y el apoyo en la prueba piloto de SAHLSA-50.

## REFERENCIAS BIBLIOGRÁFICAS

1. Parker RM. Health literacy: a challenge for American patients and their health care providers. *Health Prom Int.* 2000; 15(4): 277-283.
2. Parker RM, Ratzan S y Lurie N. Health literacy: a policy challenge for advancing high-quality health care. *Health Aff.* 2003; 22(4): 147-153.
3. Gazmararian JA, Parker RM y Baker DW. Reading skills and family planning knowledge and practices in a low-income managed-care population. *Obstet Gynecol.* 1999; 93(2): 239-244.
4. Osborne H. Health Literacy from A to Z: Practical ways to communicate your health message. Sudbury, Massachusetts: Jones and Bartlett; 2005.
5. Aguirre A, Ebrahim N y Shea J. Performance of the English and Spanish S-TOFHLA among publicly insured Medicaid and Medicare patients. *Patient Educ Couns.* 2005; 56(3): 332-339.
6. Baker D, Williams M, Parker R, Gazmararian J y Nurss J. Development of a brief test to measure functional health literacy. *Patient Educ Couns.* 1999; 38(1): 33-42.
7. Selden CR, Zorn M, Ratzan S y Parker RM (Eds). *Health Literacy. Current Bibliographies in Medicine.* 2000-1. 2002; Bethesda, MD: National Library of Medicine. Disponible en <http://www.nlm.nih.gov/archive/20061214/pubs/cbm/hliteracy.html>. [Consultado el 25 de junio de 2008].
8. Shouu-Yih D, Bender D, Ruiz R y Young C. Development of an easy-to-use Spanish health literacy test. *Health Serv Res.* 2006; 41(4): 1392-1412.
9. Zickar M. Modeling item-level data with item response theory. *Curr direct Psych Sci.* 1998; 7(4): 104-109.
10. Van der Linden, W y Hambleton, RK, eds. *Handbook of modern item response theory.* New York: Springer-Verlag.
11. Canales-Negrón R, Chiroque-Solano R, Quiliano-Terreros R, Vilchez-Román C. Aplicabilidad de la prueba de alfabetismo en salud SAHLSA-50 en pacientes mujeres del hospital Dos de Mayo. Manuscrito en proceso de arbitraje.
12. Zickar M. Modeling item-level data with item response theory. *Curr direct Psych Sci.* 1998; 7(4): 104-109.

**Correspondencia electrónica:** cvilchez@pucp.edu.pe

Fecha de recepción del artículo: 03/07/08

Fecha de aceptación del artículo: 10/09/08